DSSTox Field Definition File:

Carcinogenic Potency Database Summary Tables – All Species (CPDBAS)

(last updated 10 April 2006)

Description: Information in this file is intended to provide a minimum level of annotation to the DSSTox SDF file created for the Carcinogenic Potency Database Summary Tables – All Species (CPDBAS), obtained from the CPDB Source website: http://potency.berkeley.edu/. CPDBAS_v2a consolidated DSSTox CPDB files for rat/mouse, hamster, dog, and non-human primates. CPDBAS_v3a includes additional fields, modified fields, and new data from the CPDB website (see CPDBAS SDF Download Page and CPDBAS_LogFile for more details). For further explanation of Source-specific fields, a user is encouraged to consult the CPDB website, listed references, and documentation. Additional information is provided on the DSSTox CPDBAS SDF Download Page http://www.epa.gov/nheerl/dsstox/sdf_cpdbas.html. Some modifications in fields (and allowable contents) were made to the original CPDB Summary Tables to facilitate use of the DSSTox SDF files in relational searching applications. All modifications are documented in the ***Comments*** section of the table below.

Previously, all ***DSSTox Standard Chemical Field*** definitions were provided in the NAMEID_FieldDefFile. Since all DSSTox Structure Data Files contain the same full complement of ***DSSTox Standard Chemical Fields*** (with a few of these fields optional), users are now referred to NAMEID SDF Download Page and the central reference documentation file located at:
http://www.epa.gov/nheerl/dsstox//DSSToxAboutDSSTox/MoreonStandardChemFields/StandardChemFieldDefTable_08Dec2005.doc

The first section of the Table below lists the ***DSSTox Standard Toxicity Fields*** employed for this database, followed by the ***CPDBAS Source-Specific Fields*** containing the toxicity information particular to CPDBAS. The ***DSSTox SDF*** column lists SDF files in which the corresponding ***Field Name*** is present. All ***Units*** and ***Descriptions*** are extracted from Source reference materials unless otherwise noted. ***Allowable Values*** list allowed field entries occurring in CPDBAS, separated by slashes for exclusive entries (i.e., cannot occur with another entry) and commas or spaces for non-exclusive entries (i.e., can occur with other values). These are defined and explained in the ***Description*** section; italicized note refers to the type of entry (e.g., *Text*); the pound symbol (#) indicates that the ***Allowable Values*** entry is a number. A pound symbol followed by a list of character options (e.g., # i, m, n, v) indicates that one or more footnote characters may follow the number entry; these refer to additional information and are defined in the ***Description*** section. Multiple entries in a single field (e.g., adr cli eso) are separated by a single space in the SDF.

Information has been added to the DSSTox SDF file to alert a user of possibly redundant chemical information in the chemical structure fields, including replicate CASRN entries in the database (these corresponding to different technical grades or formulations in most cases), and to cases where 2D structures or simplified parent forms of salts or complexes have replicate entries in the database (e.g., cis and trans isomers of the same chemical, or Na and K salts of the same chemical). Two fields, **ChemicalNote** and **ChemicalReplicateCount**, contain information pertaining to replicate CASRN, parent structure, or 2D structure entries and can be used to find and count the number of unique chemical substances, parent structures, or 2D structures in the database. Note that the use of this term "replicate" in the Standard Chemical Fields pertains only to chemical information in the database that may require extra consideration before use in structure-activity modeling; replicate records in the CPDBAS always correspond to different experimental studies and biological results.

The **ToxicityNote** field is used for two primary purposes in CPDBAS. Two species listed in the Source CPDB Summary Tables (Bush Baby and Tree Shrew) have data for only a single chemical record each, where the chemical in each case is also listed in the CPDB Rat Mouse table. Rather than create a field for a single non-blank entry, we provide an entry in the **Species** field and the Bush Baby and Tree Shrew TD50 and Target Sites for the corresponding chemical record are listed in the **ToxicityNote** field. Additionally, the **ToxicityNote** field is used to document any additions of new chemical record or data modifications from CPDB(RM,HA,DG,PR)_v1 to CPDBAS_v2a and subsequent versions.

Source Website: The CPDB, from which the Summary Tables and CPDBAS are derived, is available in several formats at: http://potency.berkeley.edu/

Source Contact: Please contact Lois Swirsky Gold for questions pertaining to the content of the CPDB Summary Tables; email: cpdb@potency.berkeley.edu
Please contact DSSTox Support for questions or comments pertaining to the DSSTox CPDBAS SDF file.

Main Citations: Publications reporting use of the DSSTox SDF file for the CPDB Summary Tables are asked to list the full DSSTox file name, including date stamp, and to cite as primary reference the following citations:

http://potency.berkeley.edu/

Gold, L.S., Slone, T.H., Ames, B.N., Manley, N.B., Garfinkel, G.B., and Rohrbach, L. (1997) Carcinogenic Potency Database. In: Gold, L.S., and Zeiger, E., Eds. Handbook of Carcinogenic Potency and Genotoxicity Databases. Boca Raton, FL: CRC Press, pp. 1-605. http://potency.berkeley.edu/CRCbook.html

Gold, L.S., Manley, N.B., Slone, T.H., and Rohrbach, L. (1999) Supplement to the Carcinogenic Potency Database (CPDB): Results of animal bioassays published in the general literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environ. Health Perspect.* 107 (Suppl. 4): 527-600. http://ehpnet1.niehs.nih.gov/docs/1999/suppl-4/toc.html

Gold, L.S., Slone, T.H., Ames, B.N., Manley, N.B.,Summary Table of Chemicals in the Carcinogenic Potency Database: Results for Positivity, Potency (TD_{50}), and Target Sites: http://potency.berkeley.edu/chemicalsummary.html

SDF Usage Notes:
Each DSSTox SDF file contains a single **STRUCTURE** field. For each chemical record, the **STRUCTURE** field entry directly corresponds to the content of the **STRUCTURE_...** fields. The **STRUCTURE_Shown** field documents the relationship between what is displayed in the **STRUCTURE** field and the actual tested chemical substance, i.e. **TestSubstance_...** fields, with the latter corresponding directly to the toxicity data field entries. Commercial chemical relational database (CRD) applications may automatically insert one or more structure identifier fields upon import or export of an SDF file (e.g., Formula, FW or Mol_ID), fields that may augment or duplicate one or more of the DSSTox Standard Chemical Fields. Since the proper ordering of fields upon SDF import into most applications requires a non-blank entry in each field of the first database record, the word "blank" is inserted in each empty text field in Record 1 for this purpose; this word should be deleted from Record 1 fields after SDF import by the user is complete, particularly in the case of pure numeric fields. Users are additionally cautioned that some fields (**STRUCTURE_SMILES** and **STRUCTURE_InChI**, in particular) may exceed the 200 character limit specified in the MDL CTFiles SDF standard (see http://www.epa.gov/nheerl/dsstox/MoreonSDFs.html), and that some CRD applications may insert a line break or truncate these fields upon SDF import or export. Finally, CRD application-specific molecular header information in the SDF file is deleted in the final DSSTox SDF files; users using CRD applications requiring a molecule header upon import of the SDF can specify either **DSSTox_SID** or **DSSTox_ID_FileName**. Upon SDF import, **DSSTox_CID** can be used to identify and manage chemical structure duplicates.

As an MS Word document, the following table is best viewed onscreen using either Normal or Web Layout View in Landscape page orientation.

Field Name	***DSSTox SDF***	***Units***	***Allowable Values***	***Description***	***Comments***
			DSSTox Standard Toxicity Fields		
Study Type *(no spaces)*	All		carcinogenicity	Field is used to label all records in the database, generally with the same entry, and is designed to facilitate record identification for cross-database structure searching. Field entry refers to the main type of toxicity study for which data is represented in the database.	Field names and content are being coordinated with the public ToxML standardization effort.

Endpoint	All		TD50, Tumor Target Sites	Field is used to label all records in the database, generally with the same entry, and is designed to facilitate record identification for cross-database structure searching. Field entry refers to the type of toxicity measure represented within the database.	Field names and content are being coordinated with the public ToxML standardization effort.
Species	All		rat, mouse, hamster, dog, rhesus, cynomolgus, tree shrew, bush baby	Field is used to label all records in the database and is designed to facilitate record identification for cross-database structure searching. Field entry refers to the species of animal(s) listed in the data record and used in the toxicity study or studies.	Field names and content are being coordinated with the public ToxML standardization effort.
CPDBAS Source-Specific Fields					
Mutagenicity_SAL_CPDB	CPDBAS	*None*	positive/ negative	A chemical is classified within the CPDB as mutagenic, i.e. "positive", in the *Salmonella* assay if it was evaluated overall as either "mutagenic" or "weakly mutagenic" by Zeiger [4] or as overall "positive" by the EPA Gene-Tox Program [5,6]. All other chemicals evaluated for mutagenicity by these two sources are reported as "negative". A *blank* entry for a chemical indicates no evaluation of mutagenicity from either source. This is a summary mutagenicity determination in the CPDB Summary Table that is based on overall evaluations (not strain-specific for *Salmonella*) from two sources of overall evaluations, using the above rule.	This field is titled "Salmonella" in the original CPDB Summary Tables; symbol entries appearing in this field were converted to the following DSSTox text equivalents: "+" = positive "-" = negative "." = *blank* (no evaluation)

TD50_Rat_mg **TD50_Mouse_mg** **TD50_Hamster_ mg** **TD50_Dog_mg** **TD50_Rhesus_mg** **TD50_Cynomolgus _mg** *(no spaces)*	CPDBAS	mg/kg-body wt/ day	NP/ IA/ # i, m, n, v/	TD50 is a standardized quantitative measure of carcinogenic potency (analogous to an LD50) and is computed in the CPDB for each species/sex/tissue/tumor type for each experiment. TD50 is defined as: "that dose-rate in mg/kg body wt/day which, if administered chronically for the standard lifespan of the species, will halve the probability of remaining tumorless throughout that period" [4-6]. In the CPDB Summary Tables, a TD50 (#) is reported for a chemical in each species with a positive evaluation of carcinogenicity in at least one test. If there is only one positive test on the chemical in the species, then the most potent TD50 from that test is reported. If more than one experiment is positive, the reported TD50 is the harmonic mean of the most potent TD50 values from each positive experiment in the species (denoted with an "m" as defined below) [2,3,7-9]. If a numerical value of the TD50 is listed, it may be accompanied by one or more of the following footnote indicators (only listed in **TD50_..._mg** column): i = Intraperitoneal or intravenous injection are the only routes of administration with positive tests in the CPDB. m = More than one positive test in the species in the CPDB; in this case the reported TD50 is the harmonic mean of those positive results [7]. n = No results evaluated as positive for this species in the CPDB are statistically significant ($p<0.1$). v = Variation is greater than ten-fold among statistically significant ($p<0.1$) TD50 values from different positive experiments in the species. If a TD50 value is not reported for the listed chemical and species, the field entry will be one of the following: "NP" indicates no positive results, "IA" indicates NCI/NTP bioassays were the only available experiments and both sexes in the species were evaluated as inadequate, "*blank*" indicates no data available for that chemical and species or species-sex category. TD50 values are reported for species: rat, mouse, hamster, dog, and rhesus and cynomolgus monkeys.	These fields appear under the column headings "Harmonic mean of TD50 (mg/kg/day)" in the original CPDB Summary Tables. The field names have been shortened to simplify and lessen the probability of error in the DSSTox SDF files. Abbreviations from the original CPDB Summary Tables appearing in the TD50 fields were converted to the following DSSTox text equivalents: "-" = NP (no positive results) "I" = IA (inadequate NCI/NTP bioassays) "." = *blank* (no data) In the CPDB Summary Table for Non-Human Primates, TD50 results were reported for only a single chemical in the case of either the Tree Shrew and Bush Baby; in each case, for a chemical already present in another species CPDB Summary Table. The single data result in each case is listed in the **ToxicityNote** field with a corresponding species entry in the **Species** field, i.e. separate **TD50_...** and **TargetSites_...** fields are not present for these two species in CPDBAS.

TD50_Rat_ mmol **TD50_Mouse_ mmol** **TD50_Hamster_ mmol** *(no spaces)*	CPDBAS	mmol/kg-bodywt/day	#	**TD50_Rat_mmol** = **TD50_Rat_mg** / **MolecularWeight** etc. See definition for **TD50_Rat_mg** for details. TD50 columns in mmol/kg/bodywt/day units provided for 3 species (Rat, Mouse, Hamster). Since mg to mmol conversion requires knowledge of the molecular weight, mmol values are provided only for cases where **TestSubstance_Description** = "single chemical compound"; "defined mixture or formulation" and **STRUCTURE_Shown** = "active ingredient in formulation" , where it is assumed the mg dose reported is the active ingredient dose; "defined mixture or formulation" and **STRUCTURE_Shown** = ""representative isomer in mixture", where isomers have the same molecular weight For use in SAR modeling, these **TD50_…_mmol..** fields are offered in pure numeric form, i.e. without text footnotes provided in **TD50_Rat_mg**; however, the corresponding footnotes still apply.	

TargetSites_Rat_ Male/ **TargetSites_Rat_ Female/** **TargetSites_Rat_ BothSexes/** **TargetSites_ Mouse_Male/** **TargetSites_ Mouse_Female/** **TargetSites_ Mouse_BothSexes/** **TargetSites_ Hamster_Male/** **TargetSites_ Hamster_Female/** **TargetSites_ Hamster_ BothSexes/** **TargetSites_Dog/** **TargetSites_ Rhesus/** **TargetSites_ Cynomolgus/**	CPDBAS	*None*	adr bon cli eso ezy gal hag hmo kid lgi liv lun meo mgl mix mvc nas nrv orc ova pan per pit pre pro ski smi spl sto sub tba tes thy ubl	Tumor target sites are reported using mnemonic codes as follows: adr = adrenal gland; bon = bone; cli = clitoral gland; eso = esophagus; ezy = ear/Zymbal's gland; gal = gall bladder; hag = harderian gland; hmo = hematopoietic system; kid = kidney; lgi = large intestine; liv = liver; lun = lung; meo = mesovarium; mgl = mammary gland; mix = mixture; myc = myocardium; nas = nasal cavity (includes tissues of the nose, nasal turbinates, paranasal sinuses and trachea); nrv = nervous system; orc = oral cavity (includes tissues of the mouth, oropharynx, pharynx, and larynx); ova = ovary; pan = pancreas; per = peritoneal cavity; pit = pituitary gland; pre = preputial gland; pro = prostate; ski = skin; smi = small intestine; spl = spleen; sto = stomach; sub = subcutaneous tissue; tba = all tumor bearing animals; tes = testes; thy = thyroid gland; ubl = urinary bladder;	These fields appear under the column headings "Rat target sites", "Mouse target sites" and "Target sites" in the original CPDB Summary Tables. The field names have been separated and standardized in the DSSTox SDF files. Abbreviations from the original CPDB Summary Tables appearing in the Target site fields were converted to the following DSSTox text equivalents: "-" = NP (no positive results) "I" = IA (inadequate NCI/NTP bioassays) "." = *blank* (no data) In the CPDB Summary Table for Non-Human Primates, TD50 results were reported for only a single chemical in the case of either the Tree Shrew and Bush Baby; in each case, for a chemical already present in another species CPDB Summary Table. The single data result in each case is listed in the **ToxicityNote** field with a corresponding species entry in the **Species** field, i.e. separate **TD50_...** and **TargetSites_...** fields are not present for these two species in CPDBAS. In the CPDB Summary Tables for Rats, Mice, and Hamsters, the notation (B) is used in conjunction with a species-sex target site, e.g., "adr (B)" to indicate that a study reported experimental results only for the two sexes combined. This notation would appear for the target site under both Male and Female headings unless one of two conditions held: (1) positive results were available for the target site in the other sex of the species; or (2) negative results were reported for the other sex of the species. A "B-" would be listed alone (i.e., without target site codes) if negative results were only found in a study where the two sexes were combined. Since these results are not resolved to species-sex, we have created separate fields to accommodate this information (**TargetSites_Rat_BothSexes**, **TargetSites_Mouse_BothSexes**, **TargetSites_Hamster_BothSexes**). Non-*blank* entries in these fields only occur under the same conditions reported in the CPDB Summary Tables, i.e. if combined-sex results are available and species-sex results are unavailable for either sex. The **TargetSites_...** for dogs and non-human primate species (rhesus, cynomolgus, tree shrew, bush baby) are reported without sex specification or with sexes combined.

			ute vag vsc H NP/ IA/	ute = uterus; vag = vagina; vsc = vascular system. "H" = A mix of carcinomas of the ear duct, Zymbal's gland, oral cavity or nasal cavity were combined by single author, Maltoni, in his category "Head cancers", which he reports as induced by the chemical. "u" = NTP assigned level of evidence "positive", noting that experiments were difficult to evaluate based on Technical Report Summaries. Target sites are reported for each sex-species group with a positive result in the CPDB. Target sites are identified on the basis of a positive author's opinion for the particular site, in any experiment in the species-sex or species, using all results from both the general literature and NCI/NTP bioassays. If a chemical has two or more target sites listed, the results may be from different experiments, and a single site may be from more than one experiment, as well [2,3]. CPDB data organized by target site have been published by Gold et al. [3]. If a tumor target site code is not reported for the listed chemical and species, the field entry will be one of the following: "NP" indicates no positive results, "IA" indicates NCI/NTP bioassays were the only available experiments and were evaluated as inadequate, "*blank*" indicates no data available for that chemical and species-sex category.	In the CPDB Summary Tables, four chemicals in rats include the footnote "H" to denote a positive opinion for a consolidated site "head tumors", by only a single research group. Maltoni et al. evaluated head as a target site for a combination of carcinomas at various sites in the head (ear duct, Zymbal's gland, nasal cavity, or oral cavity). When this result has confirming results from other studies for either nas or orc target sites in the head, it is reported as nash and orch in the CPDB Summary Table. When no other positive results are reported, the result is designated "+H" in the CPDB Summary Table [4]. Both "H" and "+H" have been replaced with "H" in the CPDBAS file. This occurs either alone, or following a target site, as in "nas H" or "orc H". In the CPDB Summary Tables, the footnote indicators "L, a, and x" are listed for 3 or fewer compounds. In the DSSTox CPDBAS database file, these footnotes are deleted from the Target Sites field and the corresponding text has been incorporated into either the corresponding **ToxicityNote** field or the **NTP_TechnicalReport** field.
ActivityCategory_ SingleCellCall *(no spaces)*	CPDBAS	*None*	0 1	A conservative assignment of carcinogenic categorical activity (positive=1; negative=0) based on the presence of at least one tumor site listed for one carcinogenicity experiment cell, i.e. a particular species/sex. 0 = NP (no positive) test result for all experiments listed; 1 = one or more TD50 and tumor site listed for one or more species/sex experiment (e.g., Rat Male, Rat Female, etc)	Similar to type of conservative activity call used by the National Toxicology Program.

ActivityCategory_ MultiCellCall *(no spaces)*	CPDBAS		0/ multisite, multisex, multispecies	A more specific assignment of carcinogenic categorical activity where a positive requires more than a single cell positive response: 0 = NP (no positive) test result for all experiments listed, or only a single tumor site reported for one carcinogenicity experiment (i.e., not multisite, multisex, or multispecies result); multisite = multiple tumor sites reported in one or more experiments; multisex = male and female sexes (possibly of different species) tested positive for one or more tumor sites; multispecies = multiple species (e.g., rat, mouse, etc) tested positive at one or more tumor sites.	CPDB Summary Table tumor site codes “mix” or “H” count as a single site or experimental observation for the purposes of this activity assignment. A result reported in the CPDB Summary Table for “Both Sexes”, i.e. where the sex was either not reported or the experiment included both male and female animals of the same species, is not considered a separate result for a ”multisex” assignment (e.g., a result listed for Male Rat and Rat Both Sexes would not be considered a “multisex” result due to the uncertainties in reporting). Type of activity call used by the Food and Drug Administration (FDA), where greater weight is given to multisite, multisex or multispecies carcinogens in modeling and in carcinogenicity assessments.
ToxicityNote	As needed to provide supplementary toxicity information	*None*	*Text*	Field used to provide supplementary information pertaining to toxicity information present in chemical record.	Single entries for **TD50_...** and **TargetSites_...** values for each of two species (bush baby and tree shrew) are listed. Data availability is also noted in the **Species** field. Additionally, field is used to document any additions of new chemical record or data modifications from CPDB(RM,HA,DG,PR)_v1 to CPDBAS_v2a to subsequent versions of CPDBAS.
NTP_ TechnicalReport *(no spaces)*	As needed to provide supplementary toxicity information	*None*	*Text*	National Toxicology Program Technical Report number of study included in this CPDBAS record. Note is also included in this field if NTP study results differ from CPDB summary call to alert users to this discrepancy. The CPDB summary call can differ from the NTP call when additional literature studies meeting CPDB protocol requirements are factored into the CPDB assessment.	Users wishing to extract only the NTP records from CPDB for modeling purposes are cautioned that in a few cases the NTP study results differ from the CPDB summary call. These cases are flagged in this field.
Website_URL	As needed		*Text*	Internet full URL address for new CPDB chemical information summary pages provided on the CPDB website, indexed by chemical. Website URL was checked at time of DSSTox database publication. Please send DSSTox Error Report if website URL address no longer works or is changed.	

Additional CPDB references:

A list of citations and text files to 100 papers by the Carcinogenic Potency Project: http://potency.berkeley.edu/listofpubs.year.html

1. Gold, L.S. and Zeiger, E., Eds. (1997). Handbook of Carcinogenic Potency and Genotoxicity Databases. Boca Raton, FL: CRC Press. http://potency.berkeley.edu/CRCbook.html

2. Gold, L.S., T.H. Slone, and B. Ames (1999) Summary of Carcinogenic Potency Database by Chemical. http://potency.berkeley.edu/chemicalsummary.html

3. Gold, L.S., N.B. Manley, T.H. Slone, and J.M. Ward (2001) Compendium of chemical carcinogens by target organ: Results of chronic bioassays in rats, mice, hamsters, dogs and monkeys. *Toxicol. Pathol.* 29: 639-652. http://potency.berkeley.edu/text/ToxicolPathol.pdf

4. Zeiger, E. (1997) Genotoxicity Database. In: Gold, L.S., and Zeiger, E., Eds. Handbook of Carcinogenic Potency and Genotoxicity Databases. Boca Raton, FL: CRC Press, pp. 687-729. http://potency.berkeley.edu/CRCbook.html

5. Kier, L.E., D.J. Brusick., A.E. Auletta, E.S. Von Halle, M.M. Brown, V.F. Simmon, V. Dunkel, J. McCann, K. Mortelmans, M. Prival, T.K. Rao, and V. Ray (1986) The *Salmonella typhimurium*/mammalian microsomal assay: A report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutat. Res.* 168: 69-240.

6. Auletta, A.E., Personal communication (with L.S.Gold).

7. Peto, R., M.C. Pike, L. Bernstein, L.S. Gold, and B.N. Ames (1984) The TD_{50}: A proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. *Environ. Health Perspect.* 58: 1-8.

8. Sawyer, C., R. Peto, L. Bernstein, and M.C. Pike (1984) Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. *Biometrics* 40: 27-40.

9. Gold, L.S., T.H. Slone, and L. Bernstein (1989) Summary of carcinogenic potency (TD_{50}) and positivity for 492 rodent carcinogens in the Carcinogenic Potency Database. *Environ. Health Perspect.* 79: 259-272.